A THESIS FOR THE DEGREE OF MASTER SCIENCE

Improvement of mungbean reference genome and QTL identification for synchronous pod maturity

BY

HANEUL JEONG

AUGUST, 2017

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Improvement of mungbean reference genome and QTL identification for synchronous pod maturity

HANEUL JEONG

ABSTRACT

Mungbean (*Vigna radiata* (L.) R. Wilczek) is a good source of carbohydrate and dietary protein in South, East and Southeast Asia. However, uneven pod maturity of mungbean leads to low harvest index and more labor. In this study, we improved genome assembly of cultivated mungbean (*V. radiata* var. radiata VC1973A) by Pacbio long reads. Total of 557 scaffolds were assembled with N50 length of 5.2 Mb. Total bases of the scaffolds were 475 Mb, which was 87.5% of the estimated mungbean genome size. To anchor the scaffolds to 11 pseudochromosomes, we construct high resolution genetic map by whole genome resequencing of 187 Recombinant Inbred Lines (RILs)

derived from a cross between VC1973A and Korean landrace V2984. In addition, quantitative trait loci (QTL) analysis for synchronous maturity of pods was conducted by using single nucleotide polymorphism (SNP) markers used to construct the genetic map. To evaluate synchronous maturity of pods, we collected phenotypic data from 187 RILs. Two QTLs for synchronous maturity were found on chromosome 4 and 7 with LOD scores 2 or higher. The improved genome assembly of mungbean will facilitate genome research and molecular breeding program. Furthermore, newly identified QTLs can help breed elite cultivars with synchronous maturity leading shorter harvesting time, less labor and higher yield.

Keywords: Mungbean (*Vigna raidata*), synchronous pod maturity, genome assembly

Student number : 2015-23011

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Mungbean (*Vigna radiata* (L.) R. wilczek) is an important legume crop in Asia, particularly in south Asia, which provides dietary protein (Gupta et al., 2014). Mungbean also contains high levels of folate and iron (Keatinge et al., 2011). As a legume crop, mungbean can be useful addition in rice-based rainfed lowland cropping systems (Herrera et al., 1997). Due to its nutrition value and cost-effectiveness for the cultivation, cultivating mungbean is steadily increasing (Shanmugasundaram et al., 2009; Kim et al., 2015). However, its non-synchronous pod maturity make the average yield of mungbean low (Fernandez and Shanmugasundaram, 1988). In addition, its non-synchronous maturity of pods leads to high labor cost for harvesting (Alam Mondal et al., 2011). Although there is a lot of interests in mungbean, synchronous pod maturity is not studied well. Therefore, the development or selection of synchronous pod maturity might be a best option to increase productivity of mungbean. For the breeding in mungbean, synchronous maturity will be a primary objective that contributes in productivity and cost effective harvesting.

The objective of this study is to improve mungbean reference genome and to identify QTL (Quantitative Trait Locus) for synchronous

maturity of pods. To improve mungbean reference genome, the mungbean genome was assembled by Pacbio long reads. In addition, the scaffolds were anchored with high-resolution genetic map which constructed by re-sequencing. For identifying QTLs for synchronous pod maturity, QTL analysis was conducted by SNP (Single Nucleotide Polymorphism) used for constructing genetic map. Improved mungbean reference genome can accelerate genome research in legume crops as well as *Vigna* species. Moreover, newly identified QTLS for synchronous maturity of pods might help breed elite cultivars with synchronous maturity.

LITERATURE REVIEW

Mungbean

Mungbean (*Vigna radiata* (L) Wilczek) belong to the subgenus Ceratotropis of genus *Vigna* in the family *Fabaceae*. Mungbean is a self-pollinating diploid crop with chromosome number of 2n = 2x = 22 and the genome size is relatively small (Arumuganathan and Earle, 1991). Mungbean is cultivated mainly in India, Pakistan, Thailand, Philippines, Sri Lanka, Myanmar and Bangladesh (Raturi et al. 2012). As well as mungbean is an important grain-legume crop and a good source of dietary protein (Gulati and Jaiwal, 1994), mungbean also be able to fix atmospheric nitrogen through root rhizobial symbiosis as a legume crop (Graham and Vance, 2003). The consumptions of mungbean has been increased 22-66% from 1984 to 2006 (Shanmugasundaram et al., 2009). Besides, growing mungbean is also steadily increasing because of their cost-effectiveness for the cultivation (Kim et al., 2015). According to recent mungbean genome study (Y. J. Kang et al., 2014), 421 Mbp of the mungbean reference genome, 80% of total estimated genome of mungbean, was assembled. Comparing with other legumes, mungbean has short life-cycle and small genome size. That indicates

mungbean could be a guid plant for other genera Vigna including black gram (V. mungo), rice bean (V. umbellate) and adzuki bean (V. angularis).

Synchronous pod maturity in mungbean

In most legume crops, indeterminate growth habit has evolved more natural than human selection (Tickoo et al., 1996). Mungbean has also has indeterminate growth habit with prolonged duration for flowering and uneven pod maturity(Yeates et al., 2000; Tah and Saxena, 2009). As a result of uneven pod maturity, harvesting index (HI) and yield is low in mungbean (Bushby and Lawn, 1992; Egli and Bruening, 2006). Besides, non-synchronous pod maturity can be inefficient for the famrers because of extra harvests (Iqbal et al., 2015). Several study reports that positive effect of early and synchronous maturity on grain yield (Afzal et al., 2003; Chen et al., 2008). Because of its indeterminate growth habit, late and non-synchronous maturity and losses due to abiotic and biotic stresses, the average yield of mungbean is low (Fernandez and Shanmugasundaram, 1988).

Genome assembly

A reference genome is a nucleic acid sequence database for a representative example of a species. The reference genome is assembled from the sequencing of DNA of donors. Because there is the genetic differences between individuals in the same species, the reference genome can be a guide for representative set of genes and be used to analysis genetic variation(R. Li et al., 2009). As a result of next-generation sequencing technologies, the cost of DNA sequencing falls and more genome sequences is generated rapidly (Gnerre et al., 2011). This technologies can produce DNA sequence with ~100,000-fold lower per-base cost than a decade ago (Bentley et al., 2008; McKernan et al., 2006). For the mungbean reference genome, 421 Mbp of the mungbean cultivar VC1973A was assembled with Illumina sequencing platform (Y. J. Kang et al., 2014). However, the data produced from Illumina platform have short read-lengths and biased genome coverage leading to fragmented genome assemblies (Ferrarini et al., 2013). According to Ferrarini, the long read have the potential to resolve complex repeats and produce longer contigs with fewer gaps.

MATERIALS AND METHODS

Plant materials and phenotyping

A Recombinant Inbred Line (RIL) population was derived from a cross between cultivated mungbean VC1973A and Korean landrace V2984 in our group (Hwang et al., 2017). For phenotyping, F10 RIL population was planted at Seoul National University Experimental Farm in Suwon, Korea (37°16'12.7"N, 126°59'19.2"E) from June 25, 2016 to October 14, 2016. During that periods, average daytime was 13 hour 18 miniute and average temperature was 24.4 ℃ in Suwon, Korea. Ten plants were planted per line with 15 cm intervals. There were 70cm intervals between lines.

For QTL analysis, Height (H), Number of Nodes (NN), Number of Branches (NB), First Flowering Time (FFT) and Synchronous Maturity of Pods (SMP) were measured. The traits except FFT were measured after plants were matured. To evaluate SMP, the RILs were harvested 8 times with a week interval from August 26 to October 14. The SMP trait were calculated as the sum of the yield of the highest peak week and the yield of higher flanking week was divided by total yield.

Genetic map construction and QTL analysis

Illumina short reads generated from 187 RILs by Illumina Hiseq4000 were used for Single Nucleotide Polymorphisms (SNP) analysis. The reads were mapped to the scaffolds by using BWA v0.7.15 (H. Li, 2013). SNPs were called from the mapped reads and filtered by Samtools v1.3 (H. Li et al., 2009) (mapping quality ≥ 30, depth ≥ 5, heterozygosity ≤ 10, missing ≤ 12). The detected SNPs were used to construct genetic map by using JoinMap v4.1(Van Ooijen, n.d.). To construct robust pseudo chromosomes, another genetic map was constructed. The second genetic map was constructed by same method but with different set of SNP markers (mapping quality ≥ 30, depth ≥ 5, heterozygosity ≤ 0, missing ≤ 18).

QTL analysis was conducted with 8966 SNP markers used for genetic map. The putative QTL positions of H, NN, NB, FFT and SMP were identified with inclusive composite interval – additive mapping (ICIM-ADD) by QTL ICIMapping v.4.0.6 (Meng et al., 2015). For the significance thresholds for QTL, a logarithm of odss (LOD) score were calculated with 1000 replications of a permutation test with 99% confidence for each QTL.

.

Genome assembly

For genome assembly, mungbean genome was sequenced by two platforms, PacBio RS II and Illumina Hiseq2000. The Pacbio raw reads were error-corrected and trimmed by Canu v1.0 (Koren et al., 2017). The corrected reads were assembled into contigs by Falcon v0.3.0 (Chin et al., 2016). The contigs were scaffolded with Illumina mate-pair reads with library sizes 5, 10 and 40-kb by using SSPACE v3.0 (Boetzer et al., 2011). The scaffolds were anchored into pseudo chromosomes using two genetic maps by ALLMAPS (Tang et al., 2015). The gaps present in the superscaffolds were filled by illumina short reads by using Gapfiller v1.10 (Boetzer and Pirovano, 2012).

Genome assembly assessment

For assessing genome assembly, Core eukaryotic genes from CEGMA v2.5 (Parra, Bradnam and Korf, 2007) were mapped to the mungbean genome assembly by using blast v2.2.31 (Camacho et al., 2009). In addition, single copy ortholog gene sets of eukatyota and embryophyta selected from BUSCO (Simão et al., 2015) were mapped to the mungbean genome assembly.

Genome comparison

The new reference genome assembled with Pacbio long reads was compared with the previous reference genome to verify how much improved the new assembly was. Evenly distributed 100 SNP markers per chromosome from the previous assembly were mapped to the newly assembled chromosomes using ±100 bp flanking sequences of each marker by blast v2.2.31. Out of 1,100 markers, the markers with 90 or higher % identity were used to compare two genome assemblies.

The genes of previous mungbean assembly genomeassembly gene were aligned to *Vigna angularis* (Y. J. Kang et al., 2015), *Medicago truncatula* (Tang et al., 2014) and *Cicer arietinum* (Varshney et al., 2013) by blast v2.2.31 for genome comparison. To study synteny and collinearity between the mungbean genome and other species, the blast results were analyzed using MCScanX (Wang et al., 2012). Among total synteny blocks, those containing 20 or more genes were used to compare two species.

RESULTS

Genome assembly with long reads

Long reads of 37 Gbp were produced by Pacbio RS II platform for VC1973A (Table 1a). These reads provided a 68-fold coverage of the estimated mungbean genome size of 543 Mbp (Y. J. Kang et al., 2014). After error-correction and quality trimming, total of 15.4 Gbp Pacbio reads were assembled into 1,588 contigs with N50 length of 2.8 Mbp. These contigs were scaffolded by mate-pair reads with three different library sizes, resulting in 557 scaffolds with N50 length of 5.2 Mbp. Total bases of the scaffolds were 475 Mbp, which is 87 % of the estimated genome size (Table 2).

Table 1. Summary statistics of sequencing

(a)

Pacbio	Total reads	Total bases (bp)	Average of read length (bp)	N (%)	GC (%)	Sequencing depth (X)
	6,944,099	37,037,906,144	5333.72	0	34.78	68.21

(b)

Illumina	No. of samples	Average of total reads	Read length (bp)	Average of total bases (bp)	Average of sequencing depth (X)	Average of properly paired ratio (%)
	187	43,700,528	151	6,598,779,804	12.15	80.58

(a) Statistics of sequencing from Pacbio RS II platform for VC1973A

(b) Statistics of sequencing from Illumina Hiseq4000 platform for 187 RILs

Table 2. Summary statistics of genome assembly

	contig	scaffold	super scaffold
Statistics for contig lengths:			
Min contig length:	14	2,018	2,018
Max contig length:	12,730,532	30,296,121	73,760,808
Mean contig length:	298120.59	1102865.3	1432704.59
Standard deviation of contig length:	1004197.42	2856324.66	7764175.92
Median contig length:	9,756	23,471	9,809
N50 contig length:	2,809,390	5,182,461	47,082,489
Statistic for numbers of contigs:			
Number of contigs:	1,588	431	332
Number of contigs >= 1kb:	1,511	431	332
Number of contigs in N50:	47	24	4
Statistics for bases in the contigs:			
Number of bases in all contigs:	473,415,492	475,334,946	475,657,923
Number of bases in contigs >= 1kb:N50	473,374,614	475,334,946	475,657,923
GC Content of contigs:	33.39%	33.24%	33.27%

Genetic map construction

For SNP/INDEL analysis and high resolution genetic map construction, 187 RILs were re-sequenced. Per line, 6.6 Gbp on average, which is ~12-fold coverage of estimated mungbean genome size, were sequenced through Illumina Hiseq4000 platform (Table 1b). Total of 9,033,407 SNPs were detected by mapping the sequence read against the scaffolds. Out of the SNP markers detected, 8,966 markers were used to construct a genetic map (Figure 1a). To anchor the scaffolds to the 11 linkage groups, another genetic map was constructed with different sets of SNP markers (Table 3, Figure 1b). Using two genetic maps, scaffolds were ordered and oriented into chromosome level. Out of 10,441 markers, 10,281 markers were used to anchor scaffolds (Table 4).

Table 3. Summary of two genetic map

	Map1	Map2
Linkage Groups	11	11
Markers (unique)	8,966	1,475
Markers per Mb	21.2	4.1
N50 Scaffolds	22	22
Scaffolds	112	76
Scaffolds with 1 marker	18	11
Scaffolds with 2 markers	3	3
Scaffolds with 3 markers	5	2
Scaffolds with >=4 markers	86	60
Total bases	422,950,601 (89.0%)	359,448,445 (75.6%)

Figure 1. Genetic map of mungbean

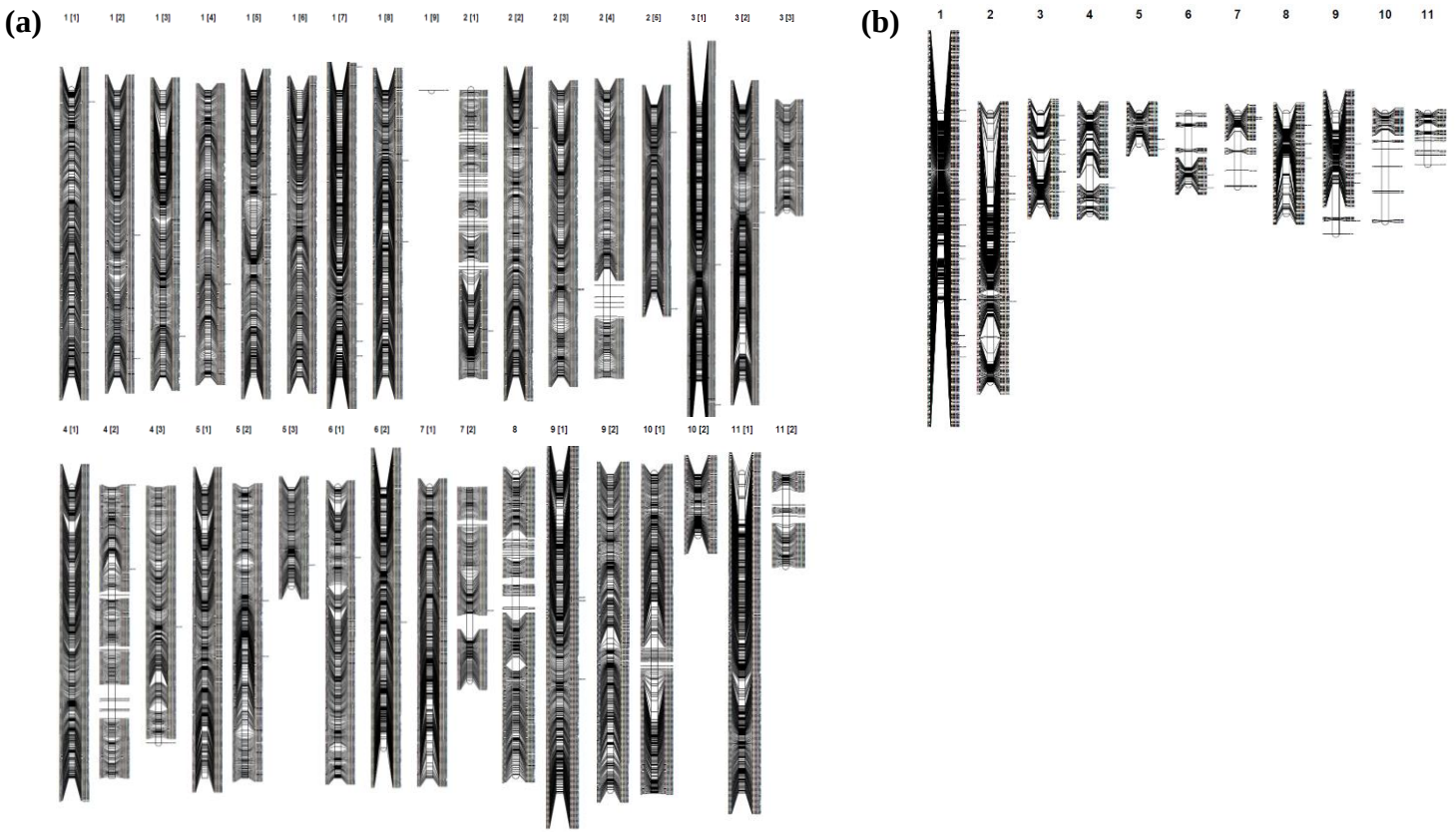


Table 4. Summary of consensus map for anchoring scaffolds

	Anchored	Oriented	Unplaced
Markers (unique)	10,270	10,155	11
Markers per Mb	24.1	26.8	0.2
N50 Scaffolds	23	22	1
Scaffolds	110	84	447
Scaffolds with 1 marker	17	0	5
Scaffolds with 2 markers	3	2	1
Scaffolds with 3 markers	4	2	0
Scaffolds with >=4 markers	86	80	1
Total bases	426,870,856 (89.8%)	379,617,302 (79.8%)	48,585,881 (10.2%)

Genome assembly assessment

For assessing genome assembly, core eukaryotic genes from CEGMA (Parra et al., 2007) were mapped to the mungbean genome assembly (Table 5). Of 248 CEGs (Core Eukayotic genes), 243 CEGs were mapped to the newly assembled chromosomes, of which 216 CEGs were mapped completely. Single copy ortholog genes of eukatoyta and embryophyta from BUSCO data sets (Simão et al., 2015) were also mapped to the mungbean genome assembly (Table 6). Out of 303 genes in eukatyotic genes, 266 genes were mapped completely and 295 genes were mapped partially to the new mungbean genome. Out of 1440 embryophyta genes, 1231 genes were mapped completely and 1303 gene were mapped partially to the mungbean genome assembly.

Table 5. Statistics of the completeness of the genome based on 248 CEGs

	Number of conserved CEGS	Percentage of conserved CEGs	Total number of CEGs including orthologs	Average number of orthologs per CEG	Percentage of detected CEGs that have more than 1 ortholog
Complete	216	87.1	364	1.69	41.67
Group 1	53	80.3	74	1.4	20.75
Group 2	47	83.93	77	1.64	42.55
Group 3	52	85.25	90	1.73	40.38
Group 4	64	98.46	123	1.92	59.38
Partial	243	97.98	457	1.88	51.44
Group 1	64	96.97	103	1.61	34.38
Group 2	55	98.21	96	1.75	47.27
Group 3	59	96.72	117	1.98	54.24
Group 4	65	100	141	2.17	69.23

Table 6. Statistics of the completeness of the genome based on gene sets from BUSCO

(a)

	Number of genes	Percentage of genes
Complete genes mapped	266	87.8%
Complete and single-copy genes	237	78.2%
Complete and duplicated genes	29	9.6%
Fragmented genes mapped	9	3.0%
Missing genes	28	9.2%
Total gene groups	303	100%

(b)

	Number of genes	Percentage of genes
Complete genes mapped	1231	85.5%
Complete and single-copy genes	1159	80.5%
Complete and duplicated genes	72	5.2%
Fragmented genes mapped	55	3.8%
Missing genes	154	10.7%
Total gene groups	1440	100%

(a) Statistics of the completeness of the genome based on eukaryote genes from BUSCO

(b) Statistics of the completeness of the genome based on embryophyta genes from BUSCO

Genome assembly improvement

One hundred markers per chromosome from the previous assembly were mapped to the newly assembled chromosomes. Out of total 1,100 markers, 959 markers were mapped to the newly assembled chromosomes with 90 or higher percent identity (Table 7). Vr03 and Vr04 of the previous reference genome assembly were combined into one chromosome, chr4 of the new reference genome assembly. From 91 and 87 markers of Vr03 and Vr04, 91 and 79 markers were mapped to chr4, respectively. In addition, 41 markers out of 87 markers from Vr08 were mapped chr4. Markers from Vr05 of the previous reference genome assembly were separated to two chromosomes, Chr10 and Chr11 of the new reference genome assembly. Out of 85 markers in Vr05, 30 and 47 markers were mapped to Chr 10 and Chr 11, respectively (Figure 2).

Of the previous mungbean genome assembly, genes from Vr03, Vr04, Vr05 and Vr08 chromosomes, which were combined or separated chromosomes in new mungbean assembly, were mapped to genes of Vigna angularis (adzuki bean), Medicago truncatula (barrelclover) and Cicer arietinum (chick pea) to find the synteny and collinearity by mapping between previous mungbean genome and other species (Table 8). To find the synteny

and collinearity, genes from previous mungbean assembly genome were mapped to genes of Vigna angularis (adzuki bean), Medicago truncatula (barrelclover) and Cicer arietinum (chick pea). Between mungbean and adzuki bean, most of the genes from Vr3 and Vr4, 43.4% and 95% genes respectively, were mapped to one chromosome, Va4 of the adzuki bean genes genome. A few Vr08 genes were also mapped to Va4. From Vr5 genes, 55.5% and 29.1% genes were mapped to Va5 and Va9, respectively (Figure 3a). In case of barrelclover and chick pea, most genes from Vr3 and Vr4 and a few genes from Vr8 were also mapped a chromosome of each species. Besides, genes of Vr5 were mapped to two different chromosomes in each species, which was consistent with the observation between previous mungbean genome and adzuki bean (Figure 3b, 3c).

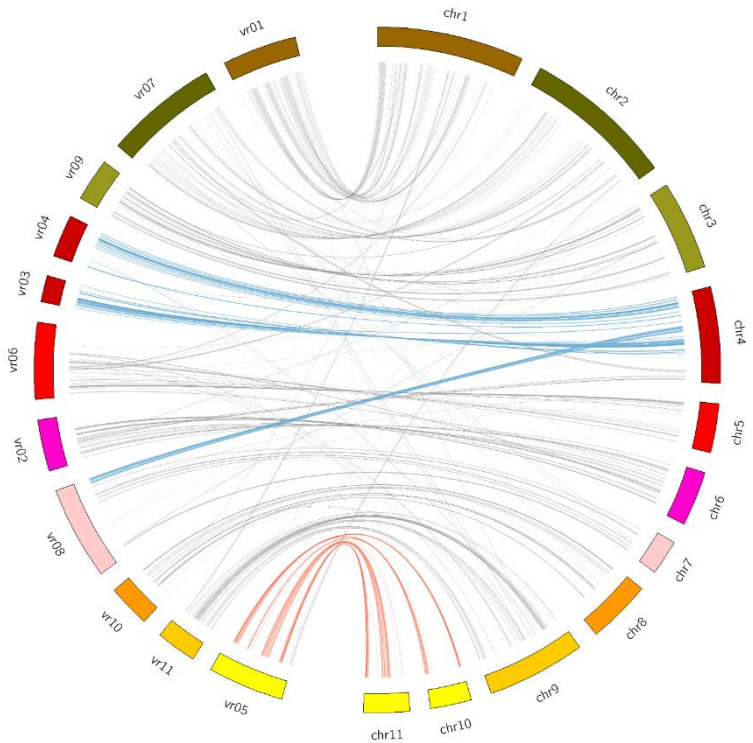
Table 7. Genome comparison between previous and new reference mungbean genome with SNP markers

Chromosome of previous reference genome	Mapped chromosome of new reference genome	Number of mapped markers	Percentage of mapped markers
Vr01	chr1	83	91.21 %
	chr2	4	4.40 %
	chr9	2	2.20 %
	chr4	1	1.10 %
	chr11	1	1.10 %
Vr02	chr6	68	76.40 %
	chr4	18	20.22 %
	chr3	3	3.37 %
Vr03	chr4	91	100.00 %
Vr04	chr4	79	90.80 %
	chr9	5	5.75 %
	chr3	2	2.30 %
	chr10	1	1.15 %
Vr05	chr11	47	55.29 %
	chr10	30	35.29 %
	chr2	8	9.41 %
Vr06	chr5	48	55.17 %
	chr6	13	14.94 %
	chr2	10	11.49 %
	chr9	9	10.34 %
	chr1	5	5.75 %
	chr3	1	1.15 %
	chr4	1	1.15 %
Vr07	chr2	62	69.66 %
	chr4	8	8.99 %

	chr1	5	5.62 %
	chr9	5	5.62 %
	chr7	3	3.37 %
	chr8	3	3.37 %
	chr3	2	2.25 %
	chr6	1	1.12 %
Vr08	chr4	41	47.13 %
	chr7	36	41.38 %
	chr3	5	5.75 %
	chr5	3	3.45 %
	chr1	1	1.15 %
	chr8	1	1.15 %
Vr09	chr3	82	93.18 %
	chr2	5	5.68 %
	chr7	1	1.14 %
Vr10	chr8	82	90.11 %
	chr9	4	4.40 %
	chr1	3	3.30 %
	chr2	1	1.10 %
	chr11	1	1.10 %
Vr11	chr9	78	90.70 %
	chr1	7	8.14 %
	chr4	1	1.16 %

Figure 2. Genome comparison between previous and new reference gnome with SNP markers

(a)



(b)

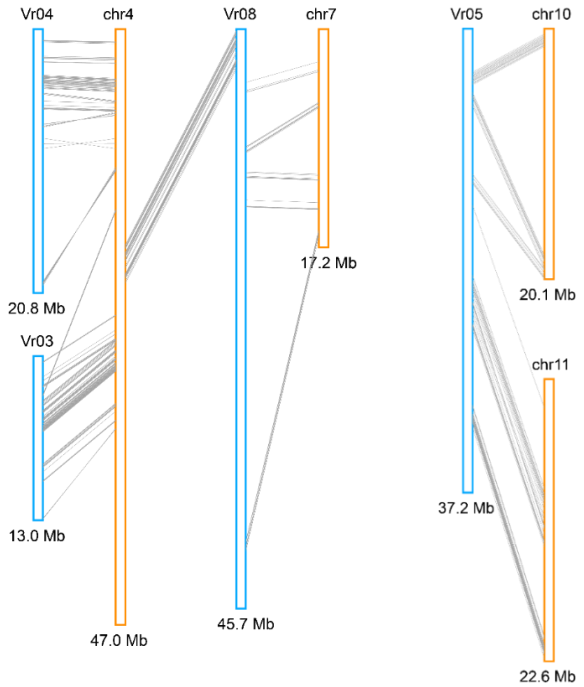
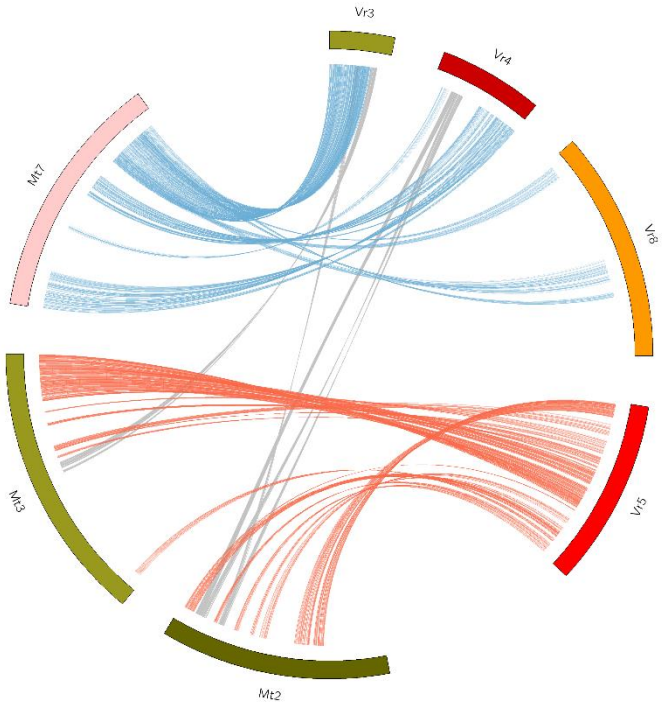


Table 8. Number of genes from synteny block between species

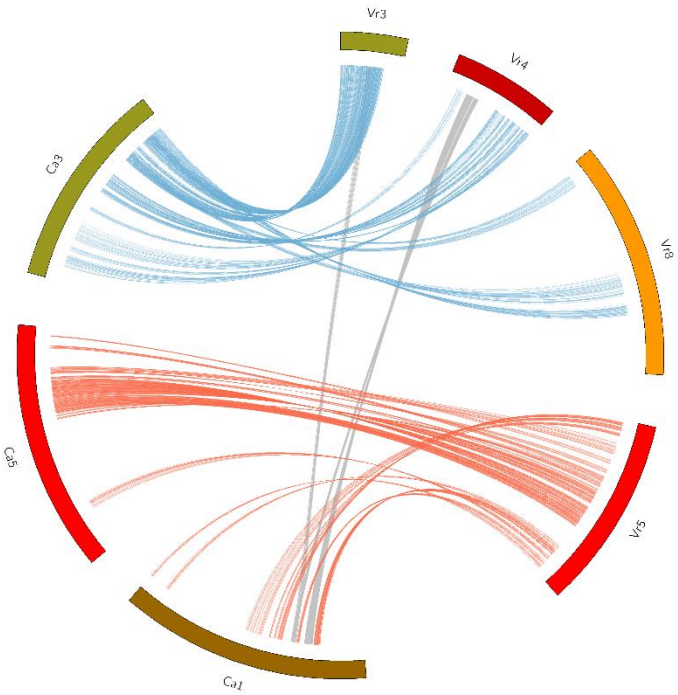
Number of genes from synteny block between species						
Mung bean	**Adzuki bean**		**Barrelclover**		**Chick pea**	
Chr	Chr	Number of genes (ratio)	Chr	Number of genes (ratio)	Chr	Number of genes (ratio)
Vr3	Va4	277 (43.42%)	Mt7	539 (57.96%)	Ca3	550 (58.14%)
	Va7	117 (18.34%)	Mt1	249 (26.77%)	Ca4	221 (23.36%)
	Va9	100 (15.67%)	Mt3	80 (8.6%)	Ca6	103 (10.89%)
	Va6	99 (15.52%)	Mt5	42 (4.52%)	Ca2	28 (2.96%)
	Va1	23 (3.61%)	Mt2	20 (2.15%)	Ca1	23 (2.43%)
	Va11	22 (3.45%)			Ca7	21 (2.22%)
Vr4	Va4	457 (95.01%)	Mt7	424 (60.49%)	Ca3	347 (55.43%)
	Va9	24 (4.99%)	Mt2	162 (23.11%)	Ca1	158 (25.24%)
			Mt4	68 (9.7%)	Ca6	84 (13.42%)
			Mt6	27 (3.85%)	Ca2	37 (5.91%)
			Mt8	20 (2.85%)		
Vr8	Va6	1613 (92.28%)	Mt1	1229 (88.29%)	Ca4	1193 (85.52%)
	Va4	58 (3.32%)	Mt7	163 (11.71%)	Ca3	202 (14.48%)
	Va3	54 (3.09%)				
	Va7	23 (1.32%)				
Vr5	Va5	626 (55.45%)	Mt3	769 (51.92%)	Ca5	736 (54.24%)
	Va9	329 (29.14%)	Mt2	447 (30.18%)	Ca1	317 (23.36%)
	Va11	92 (8.15%)	Mt4	135 (9.12%)	Ca4	143 (10.54%)
	Va1	82 (7.26%)	Mt1	130 (8.78%)	Ca7	112 (8.25%)
					Ca6	49 (3.61%)

Figure 3. Synteny block comparison between previous mungbean's gene and other species' gene

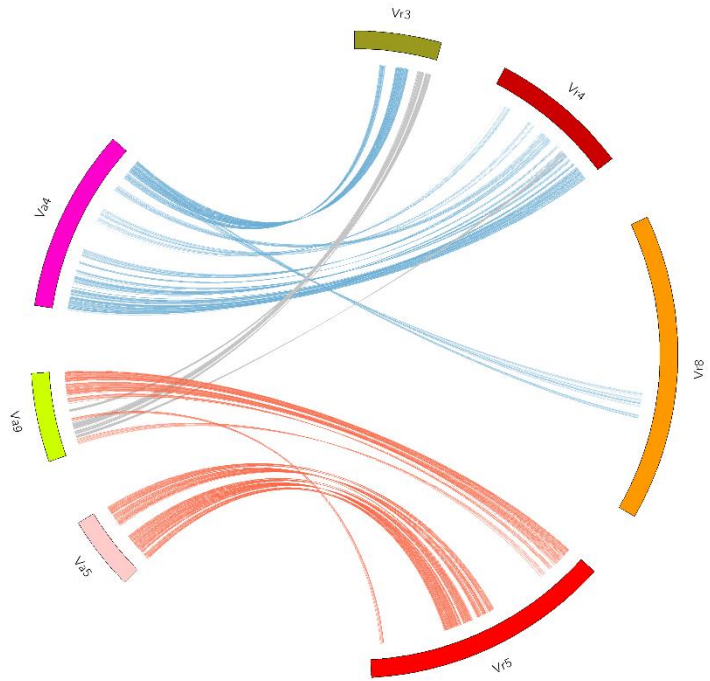
(a)



(b)



(c)



(a) synteny block comparison between mungbean and adzuki bean

(b) synteny block comparison between mungbean and chick pea

(c) synteny block comparison between mungbean and barrelclover

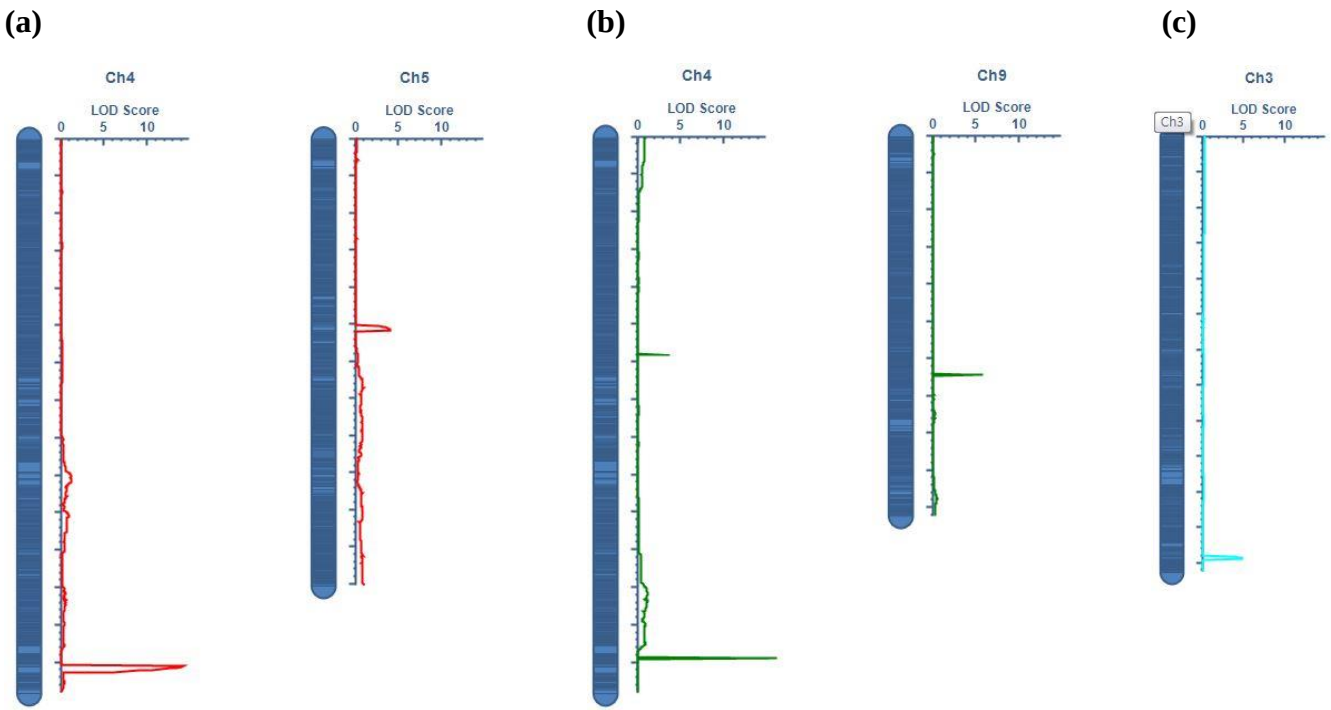
QTL analysis for agronomic trait

QTLs for H, FFT, NB, NN and SMR were identified using the ICIM-ADD method by QTL ICIMapping v4.0.6 (Meng et al., 2015) (Table 9). Each QTLs was detected with higher LOD scores than a threshold LOD scores calculated by a permutation test with 1000 replication. Two QTLs for H was detected on chr4 and chr5 with 23.9 and 6.2 percentage of variance explained (PVE), respectively (Figure 4a). For FFT traits, total of four QTLs were found and the QTLS of the two highest LOD scores were on chr4 and chr9 with 24.1 and 6.4 PVE for each (Figure 4b). In case of NB, a QTLs were observed on chr3 with 11 PVE (Figure 4c). A total of 3 QTLs for NN were detected and the two significant QTLS were on chr4 and chr11 with 20 and 6.3 PVE respectively (Figure 4d). For SMT, two QTLs were found on chr4 and chr7 with 10.3 and 6.8 PVE individually (Figure 4e).

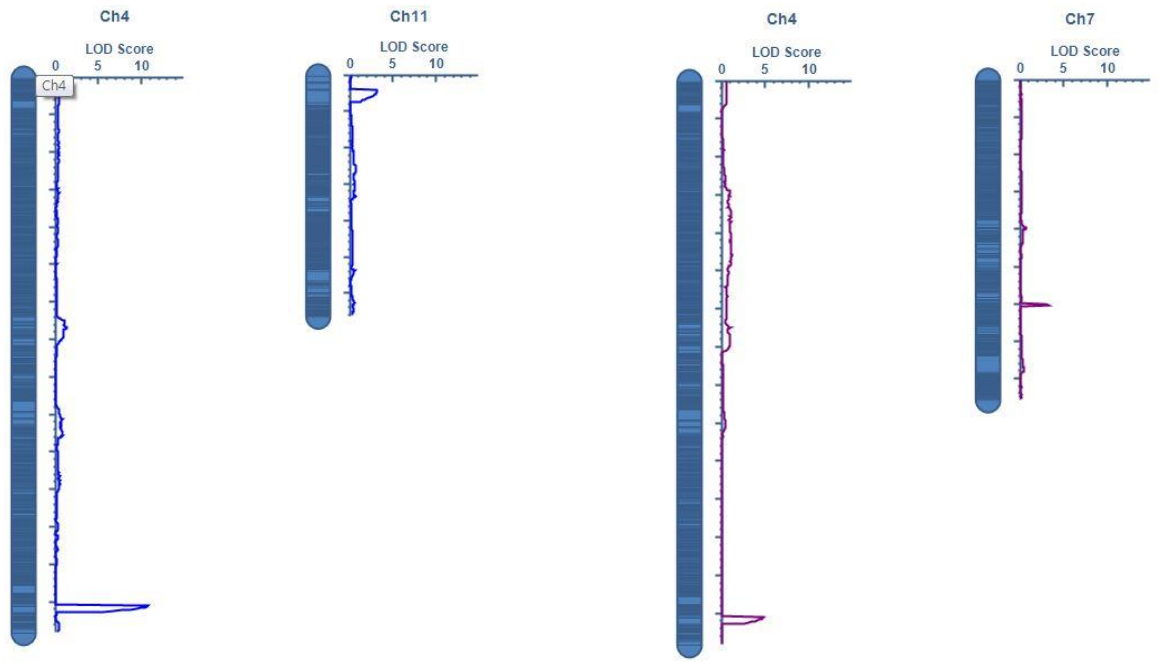
Table 9. Identified QTLs with major effects on agronomic trait

Trait	Chr	Position	Left marker	Right marker	LOD	PVE(%)	Add
H	4	704.0000	scaffold1_17425796	scaffold1_17194389	14.2334	24.9188	-5.9474
	5	258.0000	scaffold14_1478047	scaffold14_1477992	4.0474	6.2067	2.9712
FFT	4	694.0000	scaffold1_17425796	scaffold1_17430133	14.8358	24.1004	-0.9201
	6	118.0000	scaffold22_1460944	scaffold22_1565198	3.952	5.5849	-0.4422
	7	231.0000	scaffold30_4045141	scaffold30_4636408	3.1662	4.5016	0.3975
	9	322.0000	scaffold3_8752840	scaffold3_8778528	4.5092	6.4449	-0.4751
NB	3	593.0000	scaffold37_2069855	scaffold37_1716645	4.7968	11.0856	0.2686
NN	2	420.0000	scaffold21_6270826	scaffold21_6179677	3.0812	5.2382	-0.1822
	4	704.0000	scaffold1_17425796	scaffold1_17194389	10.7518	20.0198	-0.3572
	11	22.0000	scaffold4_11090118	scaffold4_8702381	3.1086	6.3403	0.2011
SMT	4	704.0000	scaffold1_17425796	scaffold1_17194389	4.7513	10.3053	0.0293
	7	301.0000	scaffold15_4109986	scaffold15_4025391	3.2413	6.8486	0.0239

Figure 4. LOD scores of identified QTLs for agronomic trait



(d) **(e)**



(a) Height (b) First Flowering Time (c) Number of branch

(d) Number of node (e) Synchronous maturity of pods

DISCUSSIONS

In this study, 431 scaffolds of 475 Mbp, 87% of total estimated mungbean genome, were assembled with N50 length of 5.2 Mbp. We constructed 11 pseudo chromosomes with the scaffolds and the SNP markers from resequencing. In other legume crop, 1,627 scaffolds of 473 Mbp covering 80 percent of total estimated common bean (P. vulgaris L..) genome was assembled with N50 length of 5 Mbp (Schmutz et al., 2014) and 3,883 scaffolds of 443 Mbp spanning 75% of total estimated adzuki bean (V. angularis var. angularis) were assembled with N50 length of 703 kbp (Y. J. Kang et al., 2015). For soybean (G. max) genome, 3,363 scaffolds of 969.6 Mb, which is 87 percent of estimated soybean genome, were assembled (Schmutz et al., 2010).

The coverage of genome and the N50 values of contigs and scaffolds were higher than previous mungbean genome assembly. In addition, we re-sequenced whole genome of mungbean while GBS (Genotype by Sequencing) was implemented on previous study to construct genetic map. For genetic map construction, total 8,966 SNP markers were used in the current study whereas 1,321 SNP markers from GBS were used on previous study.

Considering non-uniform distribution of sequencing coverage in GBS, we constructed much higher resolution genetic map than the previous one.

Putative QTLs for H, FFT, NN, NB and SMT were identified by using high resolution genetic map consisting of 8966 SNP markers. Interestingly, the most significant QTLS for H, NN and SMP were exactly same locus on chr4. Several study reports that H and NN were significantly correlated (Kang et al., 2017; Mauro-Herrera and Doust, 2016). In addition, putative flowering gene of mungbean, *Vradi03g07170*, was detected on QTLs of chr4. The light receptor phytochrome A (phyA) protein, involved in the photoperiod, was encoded in this gene (Liu et al., 2008; Watanabe et al., 2009). Furthermore, this gene was already known as candidate mungbean flowering gene (Hwang et al., 2017). In case of QTLs for SMT, its position is near with position of FFT. First flowering time was highly correlated with flowering synchrony (Ollerton and Lack, 1998). Synchronous flowering might affect pod initation, leads to synchronous maturity of pods.

In conclusion, mungbean reference genome was improved by assembling Pacbio long reads and high resolution genetic map from re-sequencing in this study. Furthermore, QTLs for agronomic trait including synchronous pod maturity were identified. Genome research in legume crops, especially in Vigna species, will be accelerated based on improved mungbean

reference genome Moreover, newly identified QTLS for synchronous maturity of pods and other agronomic traits might help elite mungbean cultivar breeding program.

REFERENCES

Afzal, M. A., Bakr, M. A., Luna, N. K., Rahman, M. M., Hamid, A., Haque, M. M., & Shanmugasundaram, S. (2003). Registration of "Barimung-5" Mungbean. *Crop Science*, *43*(6), 2304–2306.

Alam Mondal, M. M., Ali Fakir, M. S., Juraimi, A. S., Hakim, M. A., Islam, M. M., & Shamsuddoha, A. T. M. (2011). Effects of flowering behavior and pod maturity synchrony on yield of mungbean ['Vigna radiata' (L.) Wilczek]. *Australian Journal of Crop Science*, *5*(8), 945.

Arumuganathan, K., & Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, *9*(3), 208–218. https://doi.org/10.1007/BF02672069

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., … Smith, A. J. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics (Oxford, England)*, *27*(4), 578–579. https://doi.org/10.1093/bioinformatics/btq683

Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, *13*, R56. https://doi.org/10.1186/gb-2012-13-6-r56

Bushby, H. V. A., & Lawn, R. J. (1992). Accumulation and partitioning of nitrogen and dry matter by contrasting genotypes of mungbean (Vigna radiata (L.) Wilczek). *Australian Journal of Agricultural Research*, *43*(7), 1609–1628. https://doi.org/10.1071/ar9921609

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Chen, L.-R., Markhart, A. H., Shanmugasundaram, S., & Lin, T.-Y. (2008). Early developmental and stress responsive ESTs from mungbean, Vigna radiata (L.) Wilczek, seedlings. *Plant Cell Reports*, *27*(3), 535–552. https://doi.org/10.1007/s00299-007-0488-3

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., … Schatz, M. C. (2016). Phased Diploid Genome Assembly with Single Molecule Real-Time Sequencing. *bioRxiv*, 056887. https://doi.org/10.1101/056887

Egli, D. B., & Bruening, W. P. (2006). Fruit development and reproductive survival in soybean: Position and age effects. *Field Crops Research*, *98*(2), 195–202. https://doi.org/10.1016/j.fcr.2006.01.005

Fernandez, G. C. J., & Shanmugasundaram, S. (1988). The AVRDC mungbean improvement program: the past, present and future. *Mungbean. Proceedings of the Second International Symposium Held at Bangkok, Thailand, 16-20 November 1987.*, 58–70.

Ferrarini, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., … Sargent, D. J. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics*, *14*, 670. https://doi.org/10.1186/1471-2164-14-670

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., … Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*(4), 1513–1518. https://doi.org/10.1073/pnas.1017351108

Graham, P. H., & Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant Physiology*, *131*(3), 872–877. https://doi.org/10.1104/pp.017004

Gulati, A., & Jaiwal, P. K. (1994). Plant regeneration from cotyledonary node explants of mungbean (Vigna radiata (L.) Wilczek). *Plant Cell Reports*, *13*(9), 523–527. https://doi.org/10.1007/BF00232949

Gupta, S. K., Bansal, R., & Gopalakrishna, T. (2014). Development and characterization of genic SSR markers for mungbean (Vigna radiata (L.) Wilczek). *Euphytica*, *195*(2), 245–258. https://doi.org/10.1007/s10681-013-0993-0

Herrera, W. T., Garrity, D. P., & Vejpas, C. (1997). Management of Sesbania rostrata green manure crops grown prior to rainfed lowland rice on sandy soils. *Field Crops Research*, *49*(2), 259–268. https://doi.org/10.1016/S0378-4290(96)01003-9

Hwang, W. J., Ha, J., Lee, T., Jeong, H., Kim, M. Y., Kim, S. K., … Lee, S.-H. (2017). A candidate flowering gene in mungbean is homologous to a soybean Phytochrome A gene. *Euphytica*, *213*(4), 79. https://doi.org/10.1007/s10681-017-1866-8

Iqbal, J., Ahsan, M., Saleem, M., & Ali, A. (2015). Appraisal of gene action for indeterminate growth in mungbean [Vigna radiata (L.) Wilczek]. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00665

Kang, B.-K., Kim, H.-T., Choi, M.-S., Koo, S.-C., Seo, J.-H., Kim, H.-S., … Lee, and J.-D. (2017). Genetic and Environmental Variation of First

Pod Height in Soybean [<italic>Glycine max</italic> (L.) Merr.]. *Plant Breeding and Biotechnology*, *5*(1), 36–44. https://doi.org/10.9787/PBB.2017.5.1.36

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., … Lee, S.-H. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. *Nature Communications*, *5*, 5443. https://doi.org/10.1038/ncomms6443

Kang, Y. J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W. J., … Lee, S.-H. (2015). Draft genome sequence of adzuki bean, *Vigna angularis*. *Scientific Reports*, *5*, srep08069. https://doi.org/10.1038/srep08069

Keatinge, J. D. H., Easdown, W. J., Yang, R. Y., Chadha, M. L., & Shanmugasundaram, S. (2011). Overcoming chronic malnutrition in a future warming world: the key importance of mungbean and vegetable soybean. *Euphytica*, *180*(1), 129–141. https://doi.org/10.1007/s10681-011-0401-6

Kim, S. K., Nair, R. M., Lee, J., & Lee, S.-H. (2015). Genomic resources in mungbean for future breeding programs. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00626

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read

assembly via adaptive k-mer weighting and repeat separation. *bioRxiv*, 071282. https://doi.org/10.1101/071282

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [Q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., & Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Research*, *19*(6), 1124–1132. https://doi.org/10.1101/gr.088013.108

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., & Abe, J. (2008). Genetic Redundancy in Soybean Photoresponses Associated With Duplication of the Phytochrome A Gene. *Genetics*, *180*(2), 995–1007. https://doi.org/10.1534/genetics.108.092742

Mauro-Herrera, M., & Doust, A. N. (2016). Development and Genetic Control of Plant Architecture and Biomass in the Panicoid Grass,

Setaria. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0151346

McKernan, K., Blanchard, A., Kotler, L., Costa, G., & Blanchard, C. E. (2006). Reagents, methods, and libraries for bead-based sequencing. Retrieved from https://www.scienceopen.com/document?vid=70fd97f6-4268-49e8-b3ee-f013cd1de2e3

Meng, L., Li, H., Zhang, L., & Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal*, *3*(3), 269–283. https://doi.org/10.1016/j.cj.2015.01.001

Ollerton, J., & Lack, A. (1998). Relationships between Flowering Phenology, Plant Size and Reproductive Success in Lotus corniculatus (Fabaceae). *Plant Ecology*, *139*(1), 35–47.

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Raturi, A., Singh, S. K., Sharma, V., & Pathak, R. (2012). Stability and Environmental Indices Analyses for Yield Attributing Traits in Indian

Vigna radiata Genotypes under Arid Conditions. *Asian Journal of Agricultural Sciences*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=DJ2012074358

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., … Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178–183. https://doi.org/10.1038/nature08670

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., … Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008

Shanmugasundaram, S., Keatinge, J. D. H., & Hughes, J. d'A. (2009). Counting on beans: mungbean improvements in Asia. *Millions Fed: Proven Successes in Agricultural Development*, 103–108.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Tah, P. R., & Saxena, S. (2009). Induced synchrony in pod maturity in mungbean (Vigna radiata). *International Journal of Agriculture and Biology*, *11*(3), 321–324.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., … Town, C. D. (2014). An improved genome release (version Mt4.0) for the model legume Medicago truncatula. *BMC Genomics*, *15*, 312. https://doi.org/10.1186/1471-2164-15-312

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., … Lu, J. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biology*, *16*, 3. https://doi.org/10.1186/s13059-014-0573-1

Tickoo, J., L. Gajraj, R. Matho and C. Manji. (1996). Plant type in mungbean (Vigna radiata (L.) wilczek). In "Proceeding of Recent Advances in Mungbean Research" Asthana A.N. and D.H. Kim (eds.). Indian Society of Pulses Research, Kanpur, India. p. 197-213.

Van Ooijen, J. . (n.d.). JoinMap® 4.0: software for the calculation of genetic linkage maps in experimental population. Kyazma BV. Retrieved March 14, 2017, from https://www.researchgate.net/publication/288636841_JoinMapR_40_software_for_the_calculation_of_genetic_linkage_maps_in_experimental_population_Kyazma_BV

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., … Cook, D. R. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nature Biotechnology*, *31*(3), 240–246. https://doi.org/10.1038/nbt.2491

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., … Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49–e49. https://doi.org/10.1093/nar/gkr1293

Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., … Harada, K. (2009). Map-Based Cloning of the Gene Associated With the Soybean Maturity Locus *E3*. *Genetics*, *182*(4), 1251–1262. https://doi.org/10.1534/genetics.108.098772

Yeates, S. J., Lawn, R. J., & Adkins, S. W. (2000). Prediction of weather damage of mungbean seed in tropical Australia. I. Relation between seed quality, weather, and reproductive development. *Australian Journal of Agricultural Research*, *51*(5), 637–648. https://doi.org/10.1071/ar99062

ABSTRACT IN KOREAN

초록

녹두 (*Vigna radiata* (L.) R. Wilczek)는 동, 남 그리고 동남아시아에서 탄수화물과 단백질의 좋은 공급원이다. 그러나 녹두 꼬투리의 불 균일한 성숙은 낮은 수확지수와 많은 노동력을 유발한다. 본 연구의 목적은 녹두 품종인 선화녹두 (*V. radiata* var. radiata VC1973A)를 긴 Pacbio read를 이용하여 유전체 어셈블리 (assembly)를 시행하여 녹두 표준 유전체를 개선하고, 이를 이용하여 녹두 유전체에서의 동시등숙성과 관련된 양적 형질 위치를 조사하는 것이다. 녹두 표준 유전체 어셈블리 결과, 5.2 Mbp의 N50 길이를 가지는 총 557 개의 스캐폴드 (scaffold)가 어셈블리 되었고, 그 전체 길이는 녹두의 예상 유전체 길이의 87.5%인 475 Mbp 였다. 또한 스캐폴드들을 위염색체 (pseudochromosome)로 연결하기 위해서 선화녹두 (VC1973A)를 모본으로 하고 경기재래5호 (V2984)를 부본으로 하는 187개의 재조합 순계통을 이용하여 리시퀀싱 (resequencing)을 수행하여 고해상도의 유전체 지도를 완성하였다. 또한 유전체 지도를 작성하는데 이용된 단일염기 다형성 (single nucleotide polymorphism)을 이용하여 동시등숙성과 연관된 양적 형질 위치 분석을 시행하였다. 분석 결과, 로드값 (LOD score) 2 이상의

동시등숙성 양적 형질 위치를 4번 염색체와 7번 염색체에서 발견하였다. 본 연구에서 새로 개선된 녹두 표준 유전체는 앞으로 녹두의 유전체 연구와 분자 육종에 이용될 수 있을 것이다. 또한, 확인된 동시등숙성과 관련된 양적 형질 위치는 적은 노동력과 함께 수확 기간을 줄이고, 더 높은 수량을 생산하는 품종의 육종에 도움이 될 것으로 생각된다.

주요어: 녹두(*Vigna radiata*), 동시등숙성, 유전체 어셈블리

학번: 2015-23011